ACADEMIA | Letters

Artificial Intelligence, Machine Consciousness and Explanation

John Kontos, Athens University

Artificial intelligence definition

Rather than attempting to define Artificial intelligence (A.I.) as a single and consolidated discipline it might be better to consider as a set of different technologies that are easier to define individually. This set can include data mining, question answering, self-aware systems, pattern recognition, knowledge representation, automatic reasoning, deep learning, expert systems, information extraction, text mining, natural language processing, problem solving, intelligent agents, logic programming, machine learning, artificial neural networks, artificial vision, computational discovery, computational creativity. Therefore artificial "Self-aware" or "conscious" systems are the products of one of these technologies.

Some history of relevant work of mine

I have published a number of papers starting in 1970 based on software systems that were implemented by my group that solved problems of "Natural Language Processing" and particularly in the sub-area of "Natural Language Question-Answering".

I made my first steps in the Machine Consciousness field by publishing a paper [1] in 1992, in which the implementation of a self-reporting question-answering system that automatically generates explanations of its reasoning was described. I followed this line of research for more than twenty years as documented at my published papers.

Academia Letters, July 2021

Corresponding Author: John Kontos, ikontos2017@gmail.com
Citation: Kontos, J. (2021). Artificial Intelligence, Machine Consciousness and Explanation. *Academia Letters*, Article 1709. https://doi.org/10.20935/AL1709.

I published my paper on "Metagnostic Question Answering from Biomedical Texts in 2009 [2].

In 2013 we published with artist Polyxeni Kasda a paper [3] entitled "Mining and Image Anomaly Explanation with Machine Consciousness" and our collaboration led to a book described below.

Book Acupuncture

The concept of the "acupuncture of a book" occurred to me when I realized that many ideas included in the book in Greek entitled "Conscious Eye" of Polyxeni Kasda at the time of its first publication were remarkable in their foresight with almost of a Pythia-like nature. Reading the book again twenty five years after its Artificial Intelligence and the Conscious Eye first publication I spotted seventeen "acupuncture points" at which I would be able, thanks to my long research experience in A.I. to insert scientific or technological "needles" to apply at these points.

These "needles" were written in the period from February 2013 to February 2015 and are presented in our 2015 book [4] together with the English edition of the "Conscious Eye". They are based mainly but not exclusively on publications that followed the first edition of the "Conscious Eye" in Greek.

In writing my "needles" I made an effort to refer to publications as recent as possible by my research group or other researchers. Besides this state of the art references, I also have in some cases indicated the time of appearance of the earliest effort to investigate the subject hinted by the relevant paragraph of the "Conscious Eye".

The Needles

The 17 needles that constitute my contribution to the book are arranged in an order that has a loose correspondence with the chronological order of the appearance in the world of the main concepts involved.

The first Needle is a short introduction to the field of Artificial Intelligence.

In the second Needle I refer to the Turing test that was proposed in 1950 but was actually performed in 1991 with real computer programs. These events were repeated annually as part of the Loebner Prize contest. Every year the most human like program is the prize winner.

I held two dialogs with the 2012 and 2013 winner programs of the Turing test Loebner prize context and I reproduced the actual record of these conversations that promptly demon-

Corresponding Author: John Kontos, ikontos2017@gmail.com
Citation: Kontos, J. (2021). Artificial Intelligence, Machine Consciousness and Explanation. *Academia Letters*, Article 1709. https://doi.org/10.20935/AL1709.

strate the lack of consciousness of these programs. Addressing questions to such systems that presuppose consciousness for answering them promptly reveals their non-human nature.

The third Needle entitled "Introspection, "Machine Consciousness" and "Autoendoscopic Systems" traces the concept of introspection to Aristotle's works and particularly in "Prior Analytics" (4th century B.C.). In this work Aristotle is apparently the first thinker to de-scribe a mechanism of deduction by the human brain. Artificial Intelligence and the Conscious Eye

This constitutes an early introspective act of science that paves the way to Machine Consciousness that is at pre-sent a very hot subject of research. The new concepts of "Machine Consciousness" and "Autoendoscopic System" are subsequently briefly introduced.

The fourth Needle refers to the concept of Gestalt that was introduced by Koffka in 1935. The perception of a Gestalt may be considered that traverses the Unconscious- Consciousness interface. The new concept of hyper-string is a modern version of Gestalt and is similar to the loop-free finite state automaton that I used as an explanation generation mechanism. The generation of explanations may be considered as the result of introspection that in its machine form it is implemented as an artificially intelligent tracing of a mechanical parsing or reasoning process.

The fifth needle refers to conceptual or cognitive maps that were introduced by Tolman in 1948. A special case is the concept of causal maps which are based on causal relations that I have been researching since 1990.

The sixth Needle entitled "The Chair" connects the recognition of the partially occluded chair in a Picasso painting with modern Artificial or Computer Vision re-search on the detection of objects in cluttered scenes.

In the seventh Needle recent research on Artificial Vision in Technology and Art is reviewed.

The eighth Needle refers to Sartre's novel "La Nausee" published in 1938 that describes dramatically how the hero overcomes his existential "nausea" by achieving the awareness of existence. The automata I use in my research may be related to the philosophical notion of "conscious automata" that emerged in the circles of thinkers opposing the Existentialists or Sartre's followers.

The ninth Needle refers to the subject of Tinkering and its connection to learning by doing and computer programming.

The tenth needle refers to metaphors which have been studied by Aristotle many ages ago. The computer analysis of metaphors is a recent development that includes my 1980 paper. In that paper I propose an algorithm for metaphor recognition by a computer program. A very large bibliography has hence been accumulated on the subject of computational analysis and synthesis of metaphors.

Corresponding Author: John Kontos, ikontos2017@gmail.com
Citation: Kontos, J. (2021). Artificial Intelligence, Machine Consciousness and Explanation. *Academia Letters*, Article 1709. https://doi.org/10.20935/AL1709.

In the rest of the needles I refer to the subjects of Metacognitive Instruction by computer, Creativity, Artificial or Computational Creativity, Computational Scientific Discovery and Innovation, Computational Creativity in Poetry, Art and Music, Automatic Analysis of Diegesis and Automatic Synthesis of Diegesis.

Machine Consciousness and Software Synthesis

Programmers understand less and less all the operation of their programs as their complexity rises above a certain level. This is really dangerous if these programs control critical infrastructure systems like air traffic control systems, power stations and energy grids but also systems like airplanes and trains.

It is urgent then that a new kind of software engineering be developed for the implementation of computer systems that "know themselves" and can give crucial answers to the "what if" and "why" questions of their users in cases of emergency or failure. Artificial Intelligence can be of help with methods resulting from the research results in the field of Machine Consciousness. Eventually software systems supporting debugging with explanations of their failures or anomalous behaviour will be very useful.

I propose that a software system must be able to generate automatically user-friendly explanations of its answers to a user's question in order to display behaviour analogous to the behaviour of a conscious agent. In [5] some more research of mine on machine consciousness and question answering is reported and in [6] a different look at machine learning is proposed.

Self-aware computing

Trends in so-called 'self-aware computing', when a computer has some kind of sense of what is going on are current.

Such trends may lead to software that can detect whether the system is malfunctioning or whether the user is struggling with correcting an error or bug of the software.

The basic structure for a self-aware system could consist of three levels. A ground level for knowing what is happening, an object level for choosing actions, and a meta-level about monitoring and control of its actions.

Efforts to make self-aware systems have been going on for many years. An early effort was that of the 'expert systems' technology that started of over 20 years ago and aimed at acting like a human expert and generating diagnostics of its actions.

Corresponding Author: John Kontos, ikontos2017@gmail.com
Citation: Kontos, J. (2021). Artificial Intelligence, Machine Consciousness and Explanation. *Academia Letters*, Article 1709. https://doi.org/10.20935/AL1709.

XAI

The US DARPA (Defense Advanced Research Projects Agency) is encouraging the development of what it calls "Explainable Artificial Intelligence" or "XAI", which is a form of self-aware computing. This is justified because self-awareness may be inferred when a system explains to its customer or user why it is generating its output, rather than just presenting the output computed as a black box.

Software Complexity and Trust

As the complexity of software increases, the parts of their program which programmers understand are getting fewer.

Additionally software may be developed by large groups of programmers, often with poor communication between them.

This practice can be very dangerous if the software is used in critical infrastructure like air traffic control, power stations, energy grids, airplanes and trains.

This situation also means that decision makers are far less likely to trust what the system is suggesting, particularly if they don't understand how the suggestion was generated. To develop self-aware systems will require a new kind of software engineering.

One form of self-aware computing which could prove very helpful is software which can tell you when and why it is failing, and what the customer should do, rather than quietly fail and leave the customer uninformed of what happened in detail.

Examples

A drone could be programmed to avoid flying into people, trees, buildings and power lines by being aware of its position relative to other objects in its surroundings.

A tractor could be programmed to run autonomously receiving instructions from a farmer using a computer that monitors the tractor with video cameras, watching the motions of the tractor, and measuring parameters, such as engine speed, fuel levels and obstacles in its path of motion.

Corresponding Author: John Kontos, ikontos2017@gmail.com
Citation: Kontos, J. (2021). Artificial Intelligence, Machine Consciousness and Explanation. *Academia Letters*, Article 1709. https://doi.org/10.20935/AL1709.

Conclusion

The future of Artificial Intelligence will be marked by the progress in Machine Consciousness and Explainable Software.

Academia Letters, July 2021

Corresponding Author: John Kontos, ikontos2017@gmail.com
Citation: Kontos, J. (2021). Artificial Intelligence, Machine Consciousness and Explanation. *Academia Letters*, Article 1709. https://doi.org/10.20935/AL1709.

Bibliography

[1] Kontos, J. (1992). "ARISTA: Knowledge Engineering with Scientific Texts". Information and Software Technology, Vol. 34, No 9, pp 611-616.

[2] Kontos, J, et al, (2009) "Metagnostic Question Answering from Biomedical Texts", HCI2009, San Diego, USA, July 19-24.

[3] Kontos J. and Kasda P. (2013). Text Mining and Image Anomaly Explanation with Machine Consciousness. Advances in Computer Science, Vol. 2, Issue 5, No.6 , November.

[4] Kontos J. and Kasda P. (2015). "Artificial Intelligence Professor John Kontos needles Poly Kasda's Conscious Eye". Book (pp 297). Published by Notios Anemos, Athens, Greece.

[5] Kontos J. , (2016). "Machine Consciousness and Question Answering". American Journal of Data Mining and Knowledge Discovery, 1(1): 7-15.

[6] Kontos J. , (2018). "Machine Learning Methods and String Matching", 6th Global Summit on Artificial Intelligence and Neural Networks ,October 15-16, 2018 Helsinki, Finland. Theme: Harnessing the power of Artificial Intelligence.

Corresponding Author: John Kontos, ikontos2017@gmail.com
Citation: Kontos, J. (2021). Artificial Intelligence, Machine Consciousness and Explanation. *Academia Letters*, Article 1709. https://doi.org/10.20935/AL1709.